

CLAIMS

1. A method for processing job requests, comprising:

receiving a new job request by a system;

determining a completion deadline for the new job request;

estimating the length of time needed to process the new job request;

determining whether the new job would cause one or more deadlines for existing jobs in the system to be missed; and

deleting a job from the system, where said deleted job requires the greatest relative processing time when compared to other existing jobs.

2. The method of claim 1, wherein the step of determining whether the new job would cause one or more deadlines and for existing jobs in the system to be missed further comprises:

determining whether a cumulative length of time necessary to process all jobs in a job queue, including the new job request, exceeds a longest completion deadline among all jobs in said queue.

3. The method of claim 2, further comprising the step of temporarily stopping processing on a job having the longest estimated processing time if the cumulative length of time necessary to process all jobs in the queue, including the new job request, exceeds said longest completion deadline.

4. The method of claim 3, further comprising the step of resuming the processing on the job having the longest estimated processing time at a later time.

YOR920030507US1

5. The method of claim 2, further comprising the step of processing the new job request if the cumulative length of time necessary to process all jobs in the queue, including the new job request, does not exceed said longest completion deadline, and if at least one system server is idle.

6. The method of claim 2, further comprising the step of processing a job at a head of the queue after processing for another job is completed.

7. The method of claim 1, further comprising the step of routing the new job to a server in a multi-server system that can complete processing of the new job the soonest.

8. The method of claim 7, further comprising the steps of:

computing the cumulative length of time necessary to process all jobs in a queue, including the new job request, for each individual server in the system; and

identifying the server for which a cumulative length of time necessary to process all jobs in the server's queue is shortest.

9. The method of claim 1, further comprising the steps of:

estimating the total job volume directed to the system; and

determining the system capacity necessary to process at least a pre-determined percentage of the total job volume.

YOR920030507US1

10. The method of claim 9, further comprising the steps of:

determining a maximum acceptable percentage of jobs that can be lost by the system; and

deriving the job service rate necessary to substantially ensure that the percentage of jobs processed is greater than one minus the maximum acceptable percentage of jobs that can be lost.

11. The method of claim 10, further comprising the step of determining the number and speed of servers necessary to maintain the requisite job service rate.

12. The method of claim 9, further comprising the step of estimating the total job volume directed to the system from at least two separate sources.

13. The method of claim 12, further comprising the step of allocating system resources among the at least two separate sources to optimize system performance and to meet pre-specified service levels established for each of said at least two separate sources.

14. The method of claim 13, further comprising the steps of:

determining the total system capacity;

for each source, collecting job volume and deadline information;

YOR920030507US1

for each source, determining an estimated profit derived as a result of satisfying job requests from each source and the cost of satisfying the job requests; and

for each source, determining the optimal system capacity to allocate to the source to maximize the profit derived from the system performance and to process at least the pre-determined percentage of jobs.

15. The method of claim 14, further comprising the step of determining, for each source, the minimum percentage of job requests from a source that can be processed by the optimized system.

16. A method for optimizing the allocation of system resources comprising:

determining, for each source from which a system receives job requests, a minimum percentage of job requests to be processed by the system, where the system receives job requests from at least two separate sources;

allocating system resources among the least two separate sources to meet job processing obligations and to maximize the total profit of the system; and

scheduling the processing order of new job requests on-line, wherein scheduling is accomplished by:

determining a completion deadline for a new job request;

estimating the length of time needed to process the new job request by the completion deadline; and

scheduling a process time for the new job request based on the completion deadline and estimated process time as compared to completion deadlines and estimated process times for existing jobs in the system.

17. The method of claim 16, further comprising the step of:

routing the new job to a server in the system that can complete processing for the new job the soonest.

18. The method of claim 17, further comprising the steps of:

computing, for each individual server in the system, the cumulative length of time necessary to process all jobs in a queue for the server, including the new job request; and

identifying the server for which the cumulative length of time necessary to process all job's in the server's queue is shortest.

19. The method of claim of claim 17, further comprising the steps of:

assigning the new job request to the end of a server's job queue; and

determining whether the cumulative length of time necessary to process all jobs in the queue, including the new job request, exceeds a longest completion deadline among all jobs in the queue.

20. The method of claim 19, further comprising the step of temporarily stopping work on a job having the longest estimated processing time if the cumulative length of time

necessary to process all jobs in the queue, including the new job request, exceeds a longest completion deadline among all jobs in the queue.

21. A system for processing job requests, wherein the system comprises at least one server, comprising:

means for receiving a new job request to the system;

means for determining a completion deadline for the new job request;

means for estimating the length of time needed to process the new job request;

means for determining whether the new job would cause one or more deadlines and for existing jobs in the system to be missed; and

means for deleting a job from the system, where said deleted job requires the greatest relative processing time when compared to other existing jobs.

22. A computer readable medium having stored thereon a plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps of:

receiving a new job request by a system;

determining a completion deadline for the new job request;

estimating the length of time needed to process the new job request;

determining whether the new job would cause one or more deadlines for existing jobs in the system to be missed; and

deleting a job from the system, where said deleted job requires the greatest relative processing time when compared to other existing jobs.

YOR920030507US1

23. The computer readable medium of claim 22 further comprising the step of:

determining whether a cumulative length of time necessary to process all jobs in a job queue, including the new job request, exceeds a longest completion deadline among all jobs in said queue.

24. The computer readable medium of claim 23, further comprising the step of temporarily stopping processing on a job having the longest estimated processing if the cumulative length of time necessary to process all jobs in the queue, including the new job request, exceeds said longest completion deadline.

25. The computer readable medium of claim 23, further comprising the step of processing the new job request if the cumulative length of time necessary to process all jobs in the queue, including the new job request, does not exceed said cumulative completion deadline for completing all jobs in the queue, and if at least one server is idle.

26. The computer readable medium of claim 22, wherein further comprising the step of:

routing the new job to a server in the system that can complete processing of the new job the soonest.

27. The computer readable medium of claim 26, further comprising the step of:

computing, for each individual server in the system, the cumulative length of time necessary to process all jobs in a queue for the server, including the new job request; and

YOR920030507US1

identifying the server for which a cumulative length of time necessary to process all jobs in the server's queue is shortest.

28. The computer readable medium of claim 22, further comprising the steps of

estimating the total job volume directed to the system; and

determining the system capacity necessary to process at least a pre-determined percentage of the total job volume.

29. The computer readable medium of claim 28, wherein the total job volume directed to the system is the sum of jobs received from at least two separate sources.

30. The computer readable medium of claim 29, further comprising the step of:

allocating system resources among the at least two separate sources to optimize system performance and to meet pre-specified service levels established for each of said at least two separate sources.

31. The computer readable medium of claim of claim 30, further comprising the steps of:

determining the total system capacity;

for each source that sends a job to the system, collecting job volume and deadline information;

for each source that sends a job to the system, determining an estimated profit derived as a result of satisfying job requests from each source and the cost of satisfying the job requests; and

for each source that sends a job to the system, determining the optimal system capacity to allocate to the source to maximize the profit derived from the system performance and to process at least the pre-determined percentage of jobs.

32. A computer readable medium having stored thereon a plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps of:

determining, for each source from which a system receives job requests, a minimum percentage of job requests to be processed by the system, where the system receives job requests from at least two separate sources;

allocating system resources among the least two separate sources to meet job processing obligations and to maximize the total profit of the system; and

scheduling the processing order of new job requests on-line, wherein scheduling is accomplished by:

determining a completion deadline for a new job request;

estimating the length of time needed to process the new job request by the completion deadline; and

scheduling a process time for the new job request based on the completion deadline and estimated process time as compared to completion deadlines and estimated process times for existing jobs in the system.

33. The computer readable medium of claim 32, further comprising the step of:

routing the new job to a server in the system that can complete processing for the new job the soonest.

34. The computer readable medium of claim 33, further comprising the steps of:

computing, for each individual server in the system, the cumulative length of time necessary to process all jobs in a queue for the server, including the new job request; and

identifying the server for which the cumulative length of time necessary to process all job's in the server's queue is shortest.

35. The computer readable medium of claim of claim 34, further comprising the steps of:

assigning the new job request to the end of a server's job queue; and

determining whether the cumulative length of time necessary to process all jobs in the queue, including the new job request, exceeds a longest completion deadline among all jobs in the queue.